

Solutions to Module 12 Practice Problems: Basics of Statistical Inference

1. A random variable X is normally distributed with mean μ and variance 25. Suppose a researcher collected 100 observations on this random variable.

(a) What would be the distribution of the average of this researcher's observations?

By the CLT, it would be normally distributed with mean μ and variance $\frac{25}{100} = 0.25$

(b) Based on the researcher's data, what would be a 95% and a 99% confidence interval for μ ?

95% CI: $[\bar{x} - 1, \bar{x} + 1]$, 99% CI: $[\bar{x} - 1.3, \bar{x} + 1.3]$

2. Consider a random variable X . A random sample of 100 observations on X yielded $\bar{x} = 52$ and standard deviation $s = 2$.

(a) What is the point estimate for the mean of X ? What is the margin of error for this estimate?

Point estimate: 52, *Margin of error:* 0.4

(b) Find the following confidence intervals for the mean of X : (i) 95%, (ii) 99%, (iii) 90%.

(i) 95% CI: [51.6, 52.4],

(ii) 99% CI: [51.48, 52.52]

(iii) 90% CI: [51.68, 52.32]

3. During 2010-2012, Natural Resources Canada performed detailed assessments of electricity consumption in 720 Canadian homes. They found an average yearly use of 6,920 kWh per year with a standard deviation of 2210.

(a) What is the point estimate of yearly electricity use by an average Canadian home in 2012? What is the standard error of the estimate?

Point estimate: 6,920 kWh, *Standard error:* $\frac{2210}{\sqrt{720}} = 82.36$

(b) Construct a 95% confidence interval for the amount of electricity used per year by an average Canadian home.

[6755.3, 7084.7]

4. Bacteria levels in water is determined by a count of the number of bacteria colonies found in 100 ml of water. To determine the water quality in a lake, 40 samples of water were taken from different parts of a lake and the number of bacteria colonies in 100 ml of water was counted. The sample average and standard deviation were found to be 14.5 and 9.2 respectively, measured in hundreds of colonies.

(a) What is the population of study here?

The entire water in the lake. The researcher wishes to find the mean number of bacteria colonies in 100 ml of lake water.

(b) What is the point estimate of bacteria levels per 100 ml of water here? What is the standard error and the margin of error here?

Point estimate: 14.5 hundred of colonies, *Standard error:* $\frac{9.2}{\sqrt{40}} = 1.45$, *Margin of error:* 2.9

(c) Construct (i) a 95% confidence interval, and (ii) a 99% confidence interval for the mean number of colonies in 100 ml of water in the lake.

(i) 95% CI: [11.6, 17.4],

(ii) 99% CI: [10.73, 18.27]

5. A large firm conducts a survey among its employees to determine what is their daily total time for commuting to and from work. 40 employees respond. Their commute times are recorded in the first sheet of the Excel (or .ods) file “GAMES_StatInf_PracticeQuestion.xlsx” (in column B).

(a) What is the population of study here?

All the employees of the firm.

(b) Based on this data, what is the firm’s estimate of the total commute time for an average employee? What is the standard error of this estimate?

See Excel (OR .ods) solutions file in the Module.

Estimate = 39.85 minutes, Standard error = 2.95

(c) Construct a 95% confidence interval for this estimate.

[33.95, 45.75]

6. In March 2021, the Pew Research Center surveyed 3,375 U.S. Hispanic adults to understand concerns among them that they or those close to them might be deported. In the overall sample, 40% of respondents said that they were "A lot or somewhat worried" that they, a close friend or a family member could be deported. The same fraction was 28% among the 1,701 respondents who were U.S. born and 51% among the 1,674 who were foreign born.

Identify the standard error and construct 95% and 99% confidence intervals for the proportion who are worried about deportation among (i) the overall group of Hispanic adults, (ii) the group of Hispanic adults who were born in the U.S., and (iii) the group of Hispanic adults who were foreign-born.

(i) standard error = 0.0084, 95% CI: [0.383, 0.417], 99% CI: [0.378, 0.422]

(ii) standard error = 0.011, 95% CI: [0.26, 0.3], 99% CI: [0.25, 0.31]

(iii) standard error = 0.012, 95% CI: [0.486, 0.534], 99% CI: [0.479, 0.541]

7. Between September 13 - 19 2021, the Pew Research Center surveyed 10,371 U.S. adults about extreme weather and government policy. In the overall sample, 46% of respondents said that the area where they live has had an extreme weather event over the past 12 months. Among the 2,595 respondents who were Republicans or Republican-leaning, this fraction was 39%, while among the 3,300 who were Democrats or Democrat-leaning, the fraction was 52%.

The survey also asked respondents to think about the government's role when it comes to building in areas at high risk from major storms, floods and wildfires. 62% of the respondents said that they are more concerned that government will *not go far enough* in limiting new construction in high-risk areas.

(a) Construct 95% confidence intervals for the proportion who say that they have experienced an extreme weather event in the recent past among (i) the overall group of U.S. adults, (ii) those who are Republican or Republican-leaning, and (iii) those who are Democrats or Democrat-leaning. Can we say if one group's "experience" is definitely different from the other group's? In the three cases, identify the population that one is making inferences about.

(i) *Population*: all U.S. adults, standard error = 0.0049, 95% CI: [0.45, 0.47]

(ii) *Population*: all Republican or Republican-leaning U.S. adults, standard error = 0.0096, 95% CI: [0.37, 0.41]

(iii) *Population*: all Democrat or Democrat-leaning U.S. adults, standard error = 0.0087, 95% CI: [0.5, 0.54]

(b) Construct (i) a 95% confidence interval, and (ii) a 99% confidence interval for the proportion who think that government will *not go far enough* in limiting new construction in high-risk areas.

standard error = 0.0048, (i) 95% CI: [0.61, 0.63], (ii) 99% CI: [0.608, 0.633]

8. Suppose a survey is conducted to determine what fraction of university students actively trade in the stock market. In the second sheet of the Excel (or .ods) file

“GAMES_StatInf_PracticeQuestion.xlsx” are the responses (in column B) of 80 students. Column A contains their subject majors.

(a) Based on this data, what is your estimate of the fraction of university students who actively trade in the stock market. What is the margin of error? Construct a 95% confidence interval for your estimate.

(b) Among Commerce majors, what is your estimate of the fraction who actively trade? What is the margin of error? Construct a 95% confidence interval for your estimate.

(c) Among non-Commerce majors, what is your estimate of the fraction who actively trade? What is the margin of error? Construct a 95% confidence interval for your estimate.

See Excel (OR .ods) solutions file in the Module.

9. A frequent traveller was wondering if there is a difference in the hotel prices quoted on two online travel websites, A and B . So she compared the difference in prices quoted on the two websites for 8 mid-range hotels for a particular date in a particular city. The average of the difference $P_A - P_B$ was \$12, and the standard deviation was 5.

(a) What is the null and the alternate hypothesis here?

$$H_0 : P_A - P_B = 0 \quad \text{vs} \quad H_A : P_A - P_B \neq 0$$

(b) Define the test-statistic here. What is the traveller’s conclusion?

$$z = \frac{12}{5/\sqrt{8}} = 6.8. \text{ Reject } H_0 \text{ i.e. reject the hypothesis that both websites quote the same price.}$$

(c) What is the p-value of the test?

$$1 \times 10^{-12}$$

10. A chain restaurant wishes to have a standardized number of 80 fries for each order of its large fries. To test how closely this is being maintained at its restaurants, it collects samples from 10 of

its restaurants. The average number of fries in its sample was 76.1, with a standard deviation of 7.8.

(a) What is the null and the alternate hypothesis here?

Denoting μ as the mean number of fries, $H_0 : \mu = 80$ vs $H_A : \mu \neq 80$

(b) Define the test-statistic here. What is the conclusion from the test?

$z = \frac{76.1-80}{7.8/\sqrt{10}} = -1.58$. Cannot reject H_0 at 95% significance i.e. cannot reject the hypothesis that the restaurants on average are serving 80 fries in each order.

(c) What is the p-value of the test?

0.114

11. A politician wanted to know if her handling of the pandemic has affected her popularity. Prior to the pandemic, her approval rating was 62%. A poll of 40 constituents now reveal that 18 of them now approve of her, while the remaining disapprove.

(a) What is the null and the alternate hypothesis here?

Denoting μ as her approval rating, $H_0 : \mu = 0.62$ vs $H_A : \mu \neq 0.62$

(b) Define the test-statistic here. What is the politician's conclusion?

$z = \frac{0.45-0.62}{\sqrt{0.62*0.38/40}} = -2.16$. Reject H_0 at 95% significance i.e. reject the hypothesis that her handling of the pandemic has not affected her popularity.

(c) Suppose the politician really wanted to know if her approval rating has significantly decreased due to her handling of the pandemic. What would be her null and alternate hypothesis for this? What would be the conclusion?

$H_0 : \mu = 0.62$ vs $H_A : \mu \leq 0.62$

Same $z = -2.16$. Now critical value = -1.645. Again, reject H_0 at 95% significance i.e. her handling of the pandemic has lowered her popularity.

12. According to the Ontario Universities' Application Centre 2018 statistics, 15% of medical school applicants were accepted to an Ontario medical school. A coaching program advertises that 25 of the 120 students who took their program that year were successful in getting admitted into medical school, and claims this as evidence of the effectiveness of their program.

(a) What is the null and the alternate hypothesis here?

Denoting μ as the admission probability after taking the coaching program, $H_0 : \mu = 0.15$
vs $H_A : \mu \geq 0.15$

(b) Define the test-statistic and the critical value here. What is the conclusion?

$z = \frac{0.21-0.15}{\sqrt{0.21*0.79/120}} = 1.61$. Now critical value = 1.645. The z -statistic is just below this critical value. So cannot reject H_0 at 95% significance i.e. the coaching program does not make a difference.

(c) What is the p-value of the test?

0.054

(d) Is there an important underlying assumption in drawing this inference from the test?

The underlying assumption is that those who are taking the program are representative of those who are applying to medical school in general. For example, if only the weaker applicants took this coaching program, then this same conclusion cannot be drawn.

13. Let us consider CO2 emissions and GDP in Canada. The third sheet of the Excel (or .ods) file “GAMES_StatInf_PracticeQuestion.xlsx” contains data on Canadian GDP per capita and Canadian CO2 emissions per capita over the period 1970 - 2018.

Run a regression of CO2 emissions per capita (dependent variable) on GDP per capita (independent variable) for this period for Canada. Is the relationship between them statistically significant?

See Excel (OR .ods) solutions file in the Module.

Negative significant relationship at 95%.

14. Let us reconsider data on cryptocurrencies from Practice Problems in the Module on Introduction to Statistics. The fourth sheet of the Excel (or .ods) file “GAMES_StatInf_PracticeQuestion.xlsx” contains data on the closing values (i.e. the end of the day value) of Bitcoin and Ethereum over the period July 1, 2021 to December 31, 2021.

Run a regression of the Ethereum price (dependent variable) on the Bitcoin price (independent variable). Is the relationship between them statistically significant?

Positive significant relationship at 95%.

See Excel (OR .ods) solutions file in the Module.

15. Let us reconsider data on stock indices from Practice Problems in the Module on Introduction to Statistics. The fifth sheet of the Excel (or .ods) file

“GAMES_StatInf_PracticeQuestion.xlsx” contains data on the closing values (i.e. the end of the day value) of the Dow Jones Industrial Average (DOW), the NASDAQ and the Toronto Stock Exchange (TSX) over the period December 1, 2021 to January 31, 2022.

(a) Run a regression of the TSX (dependent variable) on the DOW (independent variable). Is the relationship between them statistically significant?

See Excel (OR .ods) solutions file in the Module.

Positive significant at 95%.

(b) Run a regression of the TSX (dependent variable) on the NASDAQ (independent variable). Is the relationship between them statistically significant?

See Excel (OR .ods) solutions file in the Module.

Positive significant at 95%.

(c) The last column of the sheet also contains data on the price of a Bitcoin over the period December 1 - 31, 2021. Run a regression of the price of a Bitcoin (dependent variable) on the NASDAQ (independent variable) over this period. Is the relationship between them statistically significant?

See Excel (OR .ods) solutions file in the Module.

No statistically significant relationship.

16. A researcher runs various regressions of a dependent variable Y on an independent variable X and reports the following. In which cases is the relationship statistically significant at the 95% level and in which cases is it not?

(a) p -value = 0.22, z -statistic = -1.2

Not significant

(b) z -statistic = -6.2

Significant

(c) p -value = 0.001

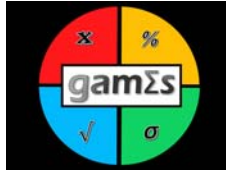
Significant

(d) $\beta = 6$, standard error (s.e.) = 4.5

z -statistic = $\frac{6}{4.5} = 1.33$; not significant

(e) $\beta = -56$, standard error (s.e.) = 7

z -statistic = $-\frac{56}{7} = -8$; significant



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) <https://creativecommons.org/licenses/by-nc-nd/4.0/> Unless otherwise noted, all content in this video was created by Catherine Pfaff, Sumon Majumdar, and Robert J. McKeown. You are free to copy and share this material in any format, but you must give appropriate credit to the authors. This project is made possible with funding by the Government of Ontario and through eCampusOntario's support of the Virtual Learning Strategy. To learn more about the Virtual Learning Strategy visit: <https://vls.ecampusontario.ca/>.